

ЛЕКЦІЯ 4. ІНДИВІДУАЛЬНІ ДІАГНОСТИЧНІ ОЗНАКИ

НАВЧАЛЬНІ ПИТАННЯ

Групування вихідних ознак

Метод потенціалів

Метод центра ваги

1. Групування вихідних ознак

Редукція ознак, що характеризують досліджуване явище, повинна відповідати деяким вимогам. Головне полягає в тому, щоб зберегти досить повний опис явища. Одержати спрощений опис, який разом з тим відбивав би найважливіші закономірності, що спостерігаються в дійсності, дозволяє методика визначення так званих індивідуальних діагностичних ознак. Дані ознаки з найбільшою вірогідністю відтворюють досліджуване явище, і при цьому їхня сукупність нечисленна. Вимоги до властивостей діагностичних ознак полягають у наступному: необхідно відкинути ознаки, що несуть однакову або досить подібну інформацію про досліджуване явище, як і ті, що залишилися; а останні у свою чергу повинні якнайкраще відбивати ті елементи, які не ввійшли в групу діагностичних ознак.

Виконання поставлених умов забезпечується підбором індивідуальних діагностичних ознак. Спочатку вихідна сукупність ознак поділяється на однорідні підмножини, тобто підмножини сильно корельованих ознак. Поділ здійснюється на основі матриці відстаней D між ознаками. Під відстанню в даному випадку будемо розуміти ступінь відмінності між ознаками. Доцільно припустити, що таку міру логічно пов'язати із коефіцієнтом кореляції між цими ознаками. Отже, елементи матриці відстаней розраховуються за формулою

$$d_{ij} = 1 - |r_{ij}|, \quad (4.1)$$

де

$$r_{ij} = \frac{1}{m} \sum_{k=1}^m z_{ki} z_{kj}. \quad (4.2)$$

z_{ki} , z_{kj} – значення стандартизованих ознак.

Якщо потім певним чином підібрати репрезентанти в кожній з виділених однорідних підмножин, то одержимо некорельовані або слабко корельовані ознаки. Це можливо зробити методом потенціалів або методом центра ваги.

1. Метод потенціалів

Сутність методу полягає в аналізі матриці відстаней і пошуку скупчень ознак навколо деяких з цих ознак. Далі для кожної вершини скупчення розраховується його потенціал. В результаті відбирається ознака з найбільшим потенціалом.

З обчислювальної точки зору алгоритм методу має наступний вигляд. На першому кроці для кожного стовпчика матриці відстаней обчислюється найменший елемент. Далі номер стовпчика і номер рядка цього елемента об'єднуються орієнтованою дугою, яка виходить з елемента з номером стовпчика. В результаті одержуються орієнтовані скупчення такого ж типу, як на рис 4.1. Далі для кожної вершини скупчення обчислюється її потенціал – кількість дуг, які входять у вершину. Тобто, фактично підраховується кількість елементів-ознак, для яких найближчою є дана ознака. Саме ця ознака і буде виступати в ролі індивідуальної від скупчення, яке являє собою однорідну підмножину ознак. Якщо в скупченні вершин з однаковим потенціалом декілька, то за індивідуальну обирають ту, для якої сума відстаней від неї до інших вершин скупчення буде найменшою.

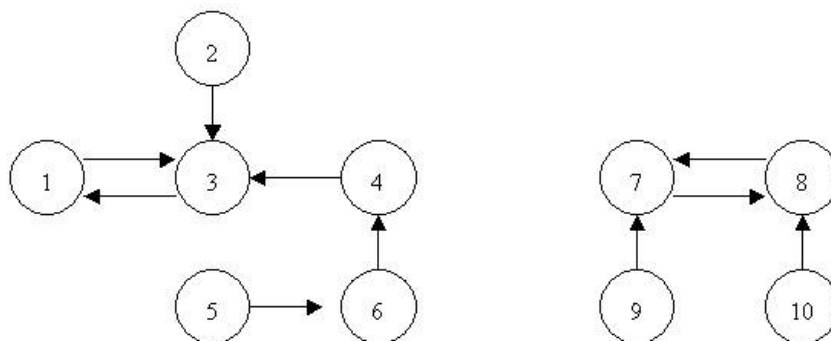


Рисунок 4.1 – Приклади орієнтованих скупчень

Для наведеного на рисунку прикладу маємо два скупчення: з 6 та 4 елементів. В першому скупченні вершина з номером 3 має найвищий потенціал – в неї входять три дуги. В другому скупченні таких вершин дві –

з номером 7 та номером 5. Отже, для них потрібно розрахувати суму відстаней до інших вершин скупчення і обрати індивідуальну за меншим значенням суми.

2. Метод центра ваги

Даний метод передбачає початкове розбиття вихідної сукупності ознак на однорідні підмножини. Це можна зробити на основі аналізу матриці відстаней між ознаками, елементи якої розраховуються за формулою (4.1). Саме розбиття можна здійснити за допомогою методу куль.

Спосіб вибору репрезентантів залежить від числа елементів, що входять в однорідну підмножину. Розрізняється три типи груп: одноелементні, двоелементні й багатоелементні.

Одноелементні групи містять такі ознаки, значення яких різко відрізняються від значень інших ознак; тому їх одразу включають у діагностичну сукупність.

Якщо кількість ознак в однорідній групі більша двох, для кожної з них розраховується сумарна відстань до інших елементів цієї ж групи:

$$d_i = \sum_{j=1}^{n_k} d_{ij}, \quad (4.3)$$

де

d_{ij} - відстань між i -тим та j -тим елементами k -ї групи;

d_j – сумарна відстань для i -того елемента k -ї групи;

n_k - число ознак у k -ї групі. У діагностичну сукупність включається та ознака, що знаходиться в "середині"

даної групи, тобто, ознака з найменшою сумою відстаней: $d_{\min i} = \min_i d_i$.

Процедура завершується вибором репрезентантів двоелементних груп. Для цього розраховуються відстані від кожного з двох елементів групи до діагностичних ознак, обраних на попередніх етапах:

$$d_i = \sum_{j=1}^s d_{ij}, \quad (4.4)$$

де $i = 1, 2$;

s - кількість відібраних діагностичних ознак.

Репрезентантом для кожної з двоелементних груп буде та ознака, відстань якої до діагностичних ознак виявиться найбільшою.

Обрані таким чином елементи володіють тією властивістю, що знаходяться поблизу "центрів ваги" груп і тому задовольняють основним вимогам, що пред'являються до діагностичних ознак.

Приклад. Завдання 4.1.

Нехай маємо дані для 8 ознак:

X_1 – трудомісткість одиниці продукції;

X_2 – питома вага робітників у складі ППП;

X_3 – питома вага придбаних виробів;

X_4 – коефіцієнт змінності обладнання;

X_5 – премії та заохочення одного робітника;

X_6 – питома вага втрат від браку;

X_7 – фондівдача;

X_8 – середньорічна чисельність ППП;

Значення ознак наведені в табл. 4.1. Потрібно побудувати індивідуальні ознаки за методами потенціалів та центра ваги.

Таблиця 4.1

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	13,26	0,23	0,78	0,40	1,37	1,23	0,23	1,45
2	10,16	0,24	0,75	0,26	1,49	1,04	0,39	1,30
3	13,72	0,19	0,68	0,40	1,44	1,80	0,43	1,37
4	12,85	0,17	0,70	0,50	1,42	0,43	0,18	1,65
5	10,63	0,23	0,62	0,40	1,35	0,88	0,15	1,91
6	9,12	0,42	0,76	0,19	1,39	0,57	0,34	0,17
7	25,83	0,31	0,73	0,25	1,16	1,82	0,38	1,94
8	23,39	0,26	0,71	0,44	1,27	1,70	0,09	1,89
9	14,68	0,49	0,69	0,17	1,16	0,84	0,14	1,94
10	10,05	0,36	0,73	0,39	1,25	0,60	0,21	2,06

Розв'язок.

А) Метод потенціалів.

Стандартизуємо ознаки, перетворивши попередньо дестимулятори в стимулятори (див розв'язок завдання

2.1). Далі обчислимо кореляційну матрицю для наведених значень ознак (табл. 4.2) і розрахуємо на її основі матрицю відстаней між ознаками (табл. 4.3). Для кожної ознаки (стовпця матриці відстаней) визначимо “найближчу” до неї ознаку (найменше стовпця значення матриці відстаней без урахування діагональних її елементів). Результат запишемо в останній рядок матриці відстаней. Ці найменші значення виділимо в табл. 4.3 кольором.

Таблиця 4.2

Кореляційна матриця							
1	0,12	0,08	-0,13	0,55	0,57	0,16	0,49
0,12	1	0,18	-0,79	-0,62	0,14	-0,13	0,41
0,08	0,18	1	-0,27	0,10	0,02	0,34	0,36
-0,13	-0,79	-0,27	1	0,32	0,12	-0,35	-0,46
0,55	-0,62	0,10	0,32	1	0,20	0,38	0,23
0,57	0,14	0,02	0,12	0,20	1	-0,26	0,34
0,16	-0,13	0,34	-0,35	0,38	-0,26	1	0,29
0,49	0,41	0,36	-0,46	0,23	0,34	0,29	1

Таблиця 4.3
Матриця відстаней

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	0	0,88	0,92	0,87	0,45	0,43	0,84	0,51
X_2	0,88	0	0,82	0,21	0,38	0,86	0,87	0,59
X_3	0,92	0,82	0	0,73	0,90	0,98	0,66	0,64
X_4	0,87	0,21	0,73	0	0,68	0,88	0,65	0,54
X_5	0,45	0,38	0,90	0,68	0	0,80	0,62	0,77
X_6	0,43	0,86	0,98	0,88	0,80	0	0,74	0,66
X_7	0,84	0,87	0,66	0,65	0,62	0,74	0	0,71
X_8	0,51	0,59	0,64	0,54	0,77	0,66	0,71	0
	X_6	X_4	X_8	X_2	X_2	X_1	X_5	X_1

В результаті одержимо два скупчення ознак (рис 4.2).

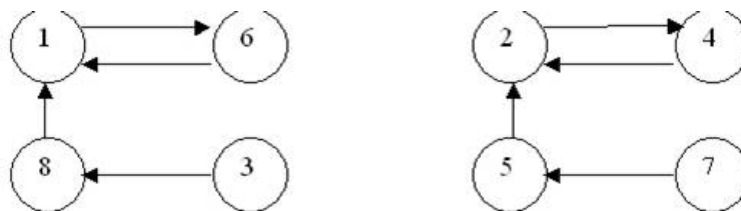


Рисунок 4.2 – Скупчення ознак для прикладу 4.1

Як видно, в кожному скупченні є вершина з найбільшим потенціалом. Для першого це вершина 1, для другого – вершина 2. Отже, в ролі індивідуальних діагностичних ознак потрібно взяти ознаки X_1 та X_2 .

Слід зазначити, що формальне застосування статистичного методу в даному випадку розділило вихідну сукупність ознак на однорідні підмножини з точки значень цих ознак та кореляційних взаємозв’язків, а не їх економічного змісту. Це є недоліком методу. Так, ознака X_8 з економічної точки зору не зовсім пов’язується з ознаками X_1 , X_3 та X_6 , з якими вона потрапила до однієї підмножини. Більш логічним було б її знаходження у другій підмножині. Тому даний метод найбільш доцільно використовувати у тому випадку, коли вихідні ознаки однорідні за економічним змістом.

Б) Метод центра ваги.

Проаналізуємо матрицю відстаней (табл. 4.3) і виділимо в ній однорідні підмножини ознак. Критичне значення радіуса кулі становить 0,64. В результаті одержимо 4 набори чотириелементних підмножин (табл. 4.4):

$\{X_1, X_5, X_6, X_8\}$

$\{X_2, X_4, X_5, X_8\}$

$\{X_1, X_2, X_5, X_7\}$

$\{X_1, X_2, X_4, X_8\}$

Таблиця 4.4 Матриця відстаней між ознаками

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	0,00	0,88	0,92	0,87	0,45	0,43	0,84	0,51
X_2	0,88	0,00	0,82	0,21	0,38	0,86	0,87	0,59
X_3	0,92	0,82	0,00	0,73	0,90	0,98	0,66	0,64
X_4	0,87	0,21	0,73	0,00	0,68	0,88	0,65	0,54
X_5	0,45	0,38	0,90	0,68	0,00	0,80	0,62	0,77
X_6	0,43	0,86	0,98	0,88	0,80	0,00	0,74	0,66
X_7	0,84	0,87	0,66	0,65	0,62	0,74	0,00	0,71
X_8	0,51	0,59	0,64	0,54	0,77	0,66	0,71	0,00
min	0,43		0,64	0,21	0,38	0,43	0,62	0,51
	4		1	3	4	2	2	4

Виходячи із змістовних міркувань, а також враховуючи, що в другому та четвертому наборах співпало по 3 елементи, зупинимось на останньому наборі ознак. Продовжуючи міркування, одержимо наступні підмножини:

Друга підмножина: $\{X_5, X_7\}$

Третя підмножина: $\{X_3\}$

Четверта підмножина: $\{X_6\}$

Таким чином, маємо дві одноелементні підмножини, одну двоелементну і одну багатоелементну. В діагностичні ознаки увійдуть X_3 та X_6 як представники одноелементних підмножин. З багатоелементної підмножини візьмемо ту ознаку, сума відстаней від якої до інших ознак підмножини найменша. Сумарні значення відстаней становлять:

X_1 – 2,26

X_2 – 1,69

X_4 – 1,72

X_8 – 1,64

Отже, репрезентантом буде ознака X_8 .

Для ознак двоелементної підмножини розрахуємо відстані від них до вже обраних репрезентантів.

X_5 – 2,48

X_7 – 2,12

Отже, репрезентантом від цієї підмножини буде ознака X_5 .

Таким чином, діагностичними ознаками будуть ознаки X_3 , X_5 , X_6 , X_8 .

Даний метод у порівнянні з попереднім дав більшу кількість діагностичних ознак.

3. Питання для самоперевірки.

1. Що являють собою індивідуальні діагностичні ознаки?
2. В чому полягає сутність методу потенціалів?
3. В чому полягає сутність методу центра ваги?
4. В чому полягає недолік використання методів побудови індивідуальних діагностичних ознак?